### Title

Model Adequacy Checking/Goodness-of-fit Testing for Behavior in Joint Dynamic Network/Behavior Models, with an Extension to Two-mode Networks




### Journal

Sociological Methods and Research, 51(4)

### ISSN

0049-1241



### Authors

Wang, C
Butts, CT
Hipp, J
et al.



### Publication Date

2022-11-01

### DOI

10.1177/0049124120914933

Peer reviewed

# Model Adequacy Checking/Goodness-of-fit Testing for Behavior in Joint Dynamic Network/Behavior Models, with an Extension to Two-mode Networks

Cheng Wang

Carter T. Butts

John R. Hipp

Cynthia M. Lakon

December 19, 2014

Post-print. Published in *Sociological Methods & Research* XX(X): XX-XX

Word count: 7,013

Word count (including references): 8,578

Running Head: "Comparing missing data strategies in SAB models"

**Model Adequacy Checking/Goodness-of-fit Testing for Behavior in Joint Dynamic Network/Behavior Models, with an Extension to Two-mode Networks**


## Abstract

The recent popularity of models that capture the dynamic coevolution of both network structure and behavior has driven the need for summary indices to assess the adequacy of these models to reproduce dynamic properties of scientific or practical importance. Whereas there are several existing indices for assessing the ability of the model to reproduce network structure over time, to date there are few indices for assessing the ability of the model to reproduce individuals' behavior patterns. Drawing on the widely used strategy of assessing model adequacy by comparing index values summarizing features of the observed data to the distribution of those index values on simulated data from the fitted model, we propose four goals that a researcher could reasonably expect of a joint structure/behavior model regarding how well it captures behavior and describe indices for assessing each of these. These reasonably simple and easily implemented indices can be used for assessing model adequacy with any dynamic network models jointly working with networks and behavior, including the stochastic actor-based models implemented within software packages such as RSien version 1.2-24. We demonstrate the use of our indices with an empirical example to show how they can be employed in practical settings, with an additional extension to modeling affiliation dynamics in two-mode networks. Key scripts are provided in the Supplemental Document (which can be found at http://smr.sagepub.com/supplemental/).



***Keywords***: model adequacy, goodness of fit (GOF), behavior dynamics, Mahalanobis distance, two-mode networks.

## Introduction

Co-evolution of (non-relational) behavior with relational structure has been a problem of long-standing interest to social scientists (Kandel 1978, 1985; Fink and Wild 1995; Dishion and Owen 2002). Although considerable theoretical investigation of coupled behavioral/relational dynamics has been conducted via simulation studies (see, e.g., Carley 1991; Leenders 1997; Mark 1998; Macy et al. 2003), research in this area has historically been limited by a paucity of available data. In recent years, the increasing availability of data including measures of both relational and *behavioral* information at multiple points in time (e.g., see West and Sweeting 1995; Harris et al. 2009; Christakis and Fowler 2007; Stelich et al. 2012; Osgood et al. 2013; Striegel et al. 2013; Purta et al. 2016) has begun to ease this limitation. At the same time, statistical techniques for fitting joint behavioral/relational models to data have begun to mature (Snijders, Steglich, and Schweinberger 2007; Steglich, Snijders, and Pearson 2010), making it possible to move beyond the qualitative insights of stylized models to the quantitative study of interdependent relational and behavioral dynamics.

This new generation of modeling techniques includes a number of distinct approaches. For example, one strategy for modeling dynamic networks is the temporal exponential random graph family (TERGM) (Robins and Pattison 2001; Hanneke et al. 2010; Krivitsky and Handcock 2014), which can be augmented with behavioral dynamics (Robins, Pattison, and Elliott 2001) and/or population dynamics (Almquist and Butts 2014) to capture evolution of vertex-level properties. Another strategy, and at present the dominant approach to studying the co-evolution of individual behavior and network structure, is Stochastic Actor-Based (SAB) modeling (Snijders, Steglich, and Schweinberger 2007; Snijders, van de Bunt, and Steglich 2010;

Snijders 2011), which models manifest network/behavioral states as arising from a latent decision process occurring in continuous time Markov processes. The SAB family is implemented with the software package RSiena (Ripley et al. 2019), making it readily available to researchers; this, and the relevance of its decision-theoretic foundation to many social network applications, has made SAB an increasingly widely used framework within the social network community (for more details on SAB modeling, see Supplemental document 1).

Regardless of the type of dynamic model being used, researchers face the same challenge of two closely related but distinctly different tasks: model assessment and model improvement. The former task involves the question of how to assess the fit of an estimated model to empirical data, with the objective of evaluating whether the fitted model could plausibly account for and/or reproduce key features of the observed data.[1] Such assessment of the fit of a model to data is often known as *model adequacy* (MA) checking in the statistical literature (see, e.g., Wasserman and Faust 1994: 365; Gelman and Meng 1996; Gelman, Meng, and Stern 1996; Gelman et al. 2003 in the context of posterior predictive assessment), or graphical *goodness of fit* (GOF) testing (Hunter et al. 2008b) by authors in the social network literature. A common motivation for this approach is the intuition that, if the model cannot even reproduce essential features (usually illustrated by *diagnostic plots* in statistical inference, e.g., see Hunter et al. 2008b; Lospinoso and Snijders 2011; Lospinoso 2012; Snijders and Steglich 2013) of the data to which it is fit, we should have little confidence in the inferences drawn from it. A second motivation is the observation (relevant in the setting in which the fitted model is proposed as a representation of the process that generated the observed data) that strong deviations of key data features from *what would typically be observed* from the fitted model requires us to presume either that the

model is not a good proxy for the generating process, or that our data is suspect. Either interpretation motivates further scrutiny.

The companion task to model assessment – model improvement – deals with the question of how to improve model performance by adding or modifying substantively meaningful parameters that reflect previously omitted or alternatively specified network/behavior evolution mechanisms. A core focus of this task lies in evaluating the relative fit of two competing models, and past network research has typically used techniques borrowed from the broader model selection literature (Hunter et al. 2008b; Lospinoso 2012). For example, in the exponential family random graph modeling (ERGM) framework, the model with smaller value of Akaike's (1973) information criterion (AIC) or Bayes information criterion (BIC) (Schwarz 1978) is typically considered a better fit (Hunter et al. 2008a, 2008b).[2] For SAB modeling, Schweinberger (2012) listed a number of forward model selection techniques – which he equated to GOF tests – to determine which parameters should be included to improve the model fit, such as the *t*-type test in the method of moment (MoM) framework (Snijders 1996), likelihood ratio tests in the maximum likelihood (ML) framework (Snijders et al. 2010), and score-type tests appropriate for both estimators (Schweinberger 2012).[3] Wang et al. (2016) introduce a different approach to assessment – held-out predictive evaluation (HOPE) – for network models that is similar to cross-validation, and which can also be used for model selection by maximizing the predicted likelihood of held-out observations; in the dynamic case, forward prediction (see e.g. Almquist and Butts, 2013) can be used for the same purpose.

While there are important differences between model adequacy and model improvement, in practice it is common for the space of potential models to be poorly specified at the outset of research (the *M*-Open paradigm, in the language of Bernardo and Smith 2007); in such cases,

model improvement requires an active process by which researchers identify weaknesses in selected models and seek alternatives that satisfy adequacy requirements while also improving fit. In this setting, adequacy checking plays a vital role in model improvement, by identifying substantively important failures in model performance and suggesting mechanisms whose inclusion or modification might result in a better model. Our work within this paper contributes to this latter dimension of model improvement. In other words, we are concerned with assessing the adequacy of the network/behavior model by inspecting the extent to which it can reproduce the observed outcome of network/behavior dynamics, whether for purposes of evaluating a single model or for identifying ways in which it might be improved.

Clearly, when testing the fit of a joint network/behavior model, it is necessary to assess how well the model captures *both* network and behavior changes. To date, many techniques are implemented in standard software packages (e.g., ergm, RSiena) for assessing *network* structure and evolution, and are readily available to analysts. By contrast, the problem of MA checking for *behavioral* evolution (or joint behavior/network evolution) has only been partially explored. Given the strong interest among researchers in disentangling the mechanisms that govern how both the network and various other behavior (e.g., delinquency) evolve over time, it is essential that studies assess the quality of proposed models for reproducing both aspects of the phenomena in question.

In the remainder of this paper, we first review our basic analytic framework for model adequacy checking for joint network/behavior models. We then introduce several reasonably simple and easily implemented indices for model adequacy checking of actor behavior in SAB or other joint network/behavior models. Finally, we apply our indices to a practical example to

illustrate their use, which includes information on friendship networks, behavior, and 2-mode networks.

## Current Strategies for Model Adequacy Checking in the Social Network Literature

With respect to model adequacy checking of network structure per se, various simulation-based techniques have been developed for assessing the extent to which a network model reproduces basic structural properties either cross-sectionally (e.g., Hunter et al. 2008b) or dynamically (e.g., Lospinoso and Snijders 2011; Lospinoso 2012; Almquist and Butts 2013). A common feature of these approaches is that they assess the micro-mechanisms of the statistical model (i.e., the parameters capturing micro decisions of network tie choice or behavioral change) by comparing the global values of key network statistics in the actual observed network to those from networks simulated based on these parameter values. This strategy is generally adopted for model adequacy checking in the network literature. For example, Hunter et al. (2008b) fitted a series of ERG models capturing a modest number of *local* mechanisms (e.g., edges, GWESP, GWD, GWDSP, and several covariates), simulated many networks from the fitted ERG models, and used diagnostic plots to illustrate whether the *global* network structure (e.g., clustering and geodesic distances) of the observed network are generally located within the predicting range of that of simulated networks. For the SAB modeling strategy, Snijders and Steglich (2013) suggested that the GOF testing is about how the fitted agency model capturing the interdependent *micro*-level processes of multiple social actors could reproduce the *macro* feature of the observed data.

It should be noted that this micro-macro linkage or local-to-global approach can be traced back to Coleman and his colleagues (e.g., Coleman 1964, 1990; Coleman, Katz, and Menzel

1966). When used in this manner, the same approach provides a sound basis for checking the adequacy of an estimated model with respect to behavior dynamics.

For the GOF testing of network properties at the global/macro level, Snijders (2003) and Mislove et al. (2007) started from a very basic family of indices – the degree distribution. Hunter et al. (2008b) introduced a group of fundamental network indices when testing the adequacy of ERG family models, including distribution of degree, edge-wise shared partners, minimum geodesic distance, and triad census,[4] and Lospinoso and Snijders (2011) and Lospinoso (2012) replaced the distribution of edgewise shared partners with Burt's constraint values when assessing the fit of SAB models.[5] Going one step further, Snijders and Steglich (2013) included several additional indices, such as size of the largest component, number of components, diameter of the largest component, the transitivity coefficient, variance of the in-degrees divided by the mean degree, variance of the out-degrees divided by the mean degree, correlation between in- and out-degrees, graph hierarchy, and least upper boundedness.

## Indices for Adequacy checking of Behavior Dynamics

Despite this burgeoning literature proposing measures for assessing the goodness of fit of various network properties, far fewer measures have been proposed for assessing the quality of the fit for individual behavior (e.g., smoking behavior, alcohol usage, etc.) being modeled along with network tie formation. We therefore focus in this paper on proposing ways to assess the quality of the fit for this portion of the model. Per Box and Draper (1987: 424), "all models are wrong, but some are useful" – our emphasis is hence not on determining whether there is *some* discrepancy between observed behavior and those predicted by the model, but rather on identifying broad patterns whose failure to replicate suggests a potentially significant weakness in the model. To assess how well a model is capturing behavior over time, we begin by

identifying some characteristics of behavioral evolution that a useful model will reproduce at the global/macro level. Although this is necessarily application-specific to some extent, there are nevertheless some desiderata that are of fairly wide applicability. We do not claim that these are the only desiderata that are of importance to researchers, but rather that they are clearly important ones upon which it is useful to begin considering model fit.

We here propose four basic goals to consider. First, one fundamental goal is that the model should reproduce the *distribution* of the behavioral variable(s) in the observed sample at each time point. Thus, we would expect a useful model to capture not just the mean level of the behavioral variables(s), but also their respective quantiles. A model that only reproduces the mean effect runs the risk of predicting either too much variance in the behavior of interest, or too little, either of which would be a serious failure on the part of the model.

Second, given the dynamic nature of the model, we would expect that a useful model should accurately capture *transitions* over time in the level of the behavioral variable(s). For example, given that a behavior of interest $z$ is modeled via $\ell$ discrete levels (from 1 to $\ell$), a $\ell \times \ell$ crosstab can be used to capture the number of persons at each level of behavior at one time point who transition to each respective level of behavior at the succeeding time point. (The row/column marginals of this table obviously constitute the information from the first desideratum.) Ideally, a useful model will produce expected transition rates that approximate those of the observed data. A model that only captures the static distribution at particular waves but fails to capture the individual-level change in the behavior is not appropriately capturing the dynamic nature of the data, and therefore would be unsatisfactory.

A third goal for a useful model is that it should generate networks in which the behavior patterns align with key statistics (e.g., vertex-level indices) of the social network. For example,

in-degree and out-degree are elementary indices associated with a range of social processes, and a useful model should ideally be able to reproduce the extent to which the behavioral variables are related to these two network statistics. Thus, we would expect e.g. that the model should be able to reproduce the average number of out-going ties for all persons at a given level of the behavior, and that it would be able to reproduce the average number of in-coming ties for all persons at a given level of the behavior. This would be evidence that the model is appropriately capturing the symbiotic relationship between network change and behavior change for individuals at specific locations in the network.

Finally, a fourth goal is that a useful model should be able to capture the degree of behavior clustering within the network. Thus, we are not simply interested in the dynamic evolution of the network and behavior at the same time, but how they each inform the other. Given that dynamic network/behavior models are typically trying to infer the extent to which social influence (the tendency to changes behavior to match that of one's friends) and social selection (the tendency to become friends with those at similar levels of behavior) lead to networks with clustered behavior based on friendship ties, it is reasonable to expect the model to reproduce such clustering.

Although these four desiderata are jointly important when assessing the fit of the behavior dynamics in a model of the dynamic co-evolution of both network structure and other behavior, we note that e.g. in the SAB modeling framework only the first has been included in existing software (Lospinoso and Snijders 2011; Lospinoso 2012; Ripley et al. 2019). Furthermore, we are aware of no published studies that have assessed the other three desiderata when assessing the fit of their models. Thus one of our goals is to develop indices of model

adequacy or goodness of fit based on these four desiderata, and to demonstrate the use of these indices in a real-world network/behavior data set.

Based on the four desiderata, we next describe some reasonably simple and easy indices for model adequacy checking of behavior dynamics at the global/macro level, which can be categorized into three groups: 1) behavior statistics, 2) behavior-based network statistics, and 3) network-based behavior similarity. These indices can be used with any joint network/behavior model family, including SAB modeling, and they can be used as the basis for omnibus discrepancy measures for goodness of fit as well as application-specific adequacy evaluation.

### *Behavior statistics*

***Behavior distribution***. One goodness of fit assessment is to compare the behavior distribution in the data to that generated by the model. Such an approach was introduced Lospinoso and Snijders (2011) and is currently provided in the RSiena software package (Ripley et al. 2019). Given a behavior of interest ($z$) modeled as ordinal via $\ell$ levels (from 1 to $\ell$) at each discrete time point, the vector of statistics for the behavior distribution is thus

$$A_{z(c,t)} = A_{z(1,t)}, A_{z(2,t)}, \ldots, A_{z(\ell,t).} \quad (1)$$

where $A_{z(c,t)}$ is a vector capturing the number of individuals adopting the behavior level $c$ at time $t$ with a length (i.e., number of elements) of $\ell$. Although current implementations of software such as RSiena cannot model continuous behavior measures (only ordinal measures), if future software develops the capability to model continuous measures, our approach described here could be utilized by binning the continuous measure after estimating the model and then performing this test. For example, it would be an argument for using the quantiles of the distribution as statistics, instead of the level counts. The former generalizes immediately to the continuous case.

***Behavior transition***. A second goal we mentioned above was to capture behavioral transition. That is, does the model capture the proportion of the sample transitioning from one level of the behavior to another by the next time point? We define this as a measure of *behavioral transition*, a notion used extensively by e.g. Coleman (1964). Given a behavior of interest *z*, modeled via $\ell$ levels (from 1 to $\ell$) at each observed time point, the behavioral transition vector of statistics is thus

$$A_{z(c,t-1)\rightarrow z(c,t)} = A_{z(1,t-1)\rightarrow z(1,t)}, A_{z(1,t-1)\rightarrow z(2,t)}, \ldots, A_{z(\ell,t-1)\rightarrow z(\ell-1,t)}, A_{z(\ell,t-1)\rightarrow z(\ell,t)} \quad (2)$$

with a length (i.e., number of elements) of $\ell^2$. The arrow ($\rightarrow$) refers to the transition, hence $A_{z(1,t-1)\rightarrow z(1,t)}$ refers to the number of cases in which the individual remains at the lowest behavioral value (1) at both time points, and $A_{z(1,t-1)\rightarrow z(2,t)}$ refers to the number of individuals that transition from the lowest value (1) to the next lowest value (2) over the two time points, and so on and so forth.

***Level of behavior change***. The measure of *level of behavior change* assesses how well the model captures the extent of change in the behavior for all individuals in the sample by the next time point. Given a behavior of interest *z* is modeled via $\ell$ levels (from 1 to $\ell$) at each observed time point, behavior change vector of statistics is thus

$$A_{z(\Delta c,t-1\rightarrow t)} = A_{z(-\ell+1,t-1\rightarrow t)}, A_{z(-\ell+2,t-1\rightarrow t)}, \ldots, A_{z(0,t-1\rightarrow t)}, \ldots, A_{z(\ell-2,t-1\rightarrow t)}, A_{z(\ell-1,t-1\rightarrow t)} \quad (3)$$

where $A_{z(\Delta c,t-1\rightarrow t)}$ is a vector capturing the change score $\Delta c$ across two time points with a length (i.e., number of elements) of $2\ell$-1. Here the first element $A_{z(-l+1,t-1\rightarrow t)}$ represents the number of individuals who decreased the maximum value (-$\ell$+1) on the behavior scale, the second element represents the number of individuals who decreased one less than the maximum value on the behavior scale (-$\ell$+2), etc. For example, if the behavior is measured with an ordinal scale from 1 to 3, then the smallest value in the vector would be -2 (as some people potentially could change

from the highest value to the lowest value on the behavior between the two time points), whereas the largest value would be 2 (as some people potentially could change from the lowest to the highest value on the behavior). And the number of persons whose change in the behavior is captured by the intervening integer values would be contained in this vector.

### *Behavior-based network statistics*

***Average out-degree by behavior***. Another goal for a model is being able to reproduce the distribution of behavior given certain values of network statistics. For example, to what extent does the model capture the out-degree distribution in the sample for each level of behavior? This is a measure of *out-degree by behavior*. Given a behavior of interest *z* is modeled via $\ell$ levels (from 1 to $\ell$) at each discrete time point, for average out-degree $\bar{x}_{i+}$ by behavior the vector of statistics is thus

$$A_{\bar{x}_{i+}|z(c,t)} = A_{\bar{x}_{i+}|z(1,t)}, A_{\bar{x}_{i+}|z(2,t)}, \ldots, A_{\bar{x}_{i+}|z(\ell,t)} \quad (4)$$

with a length (i.e., number of elements) of $\ell$. Each element represents the average out-degree ($\bar{x}_{i+}$) conditional on (|) a particular value of the behavior. Here the element $A_{\bar{x}_{i+}|z(c,t)}$ is the average out-degree for individuals adopting behavior level *c* at time *t*.

***Average in-degree by Behavior***. Another network statistic of interest is in-degree. Thus, we might wish to assess the model's ability to reproduce the in-degree distribution in the sample for each level of the behavior. This is a measure of *in-degree by behavior*. Given a behavior of interest *z* is modeled via $\ell$ levels (from 1 to $\ell$) at each discrete time point, for average in-degree $\bar{x}_{+i}$ by behavior the vector of statistics is thus

$$A_{\bar{x}_{+i}|z(c,t)} = A_{\bar{x}_{+i}|z(1,t)}, A_{\bar{x}_{+i}|z(2,t)}, \ldots, A_{\bar{x}_{+i}|z(\ell,t)} \quad (5)$$

with a length (i.e., number of elements) of $\ell$. Here the element $A_{\bar{x}_{+i}|z(c,t)}$ is the average in-degree for individuals adopting behavior level *c* at time *t*.

*Network-based behavior similarity*

***Edgewise homophily***. A fourth goal of the model is to reproduce the degree of behavior clustering within the network. One way to assess this is through the technique of edgewise homophily (EH) (Lospinoso and Snijders 2011, and Lospinoso 2012).[6] This index assesses the behavior similarity among edges. The original function of edgewise homophily (Lospinoso and Snijders 2011; Lospinoso 2012) is the sum of possible dyadic ($x_{ij}$) isomorphisms (mutual, asymmetric, and null) multiplied by the behavior similarity function (**S**). Note that the behavior similarity function (**S)** is general, and thus needs to be defined for a specific instance. Thus,

$$A_{EH(t)} = \sum_{i<j} \begin{pmatrix} x_{ij(t)}x_{ji(t)} \\ x_{ij(t)}(1 - x_{ji(t)}) \\ (1 - x_{ij(t)})(1 - x_{ji(t)}) \end{pmatrix} \boldsymbol{S}_{(t)}. \quad (6)$$

Here we use the similarity score (Ripley et al. 2019: 187) to calculate the behavior similarity function in equation (6). And we also normalize the original measure with the denominator in equation (7) – the number of edges – for two reasons: 1) to avoid conflation with the number of edges, and 2) the number of edges in the observed and simulated data might not be the same.

$$A_{EH(t)} = \frac{\sum_{i<j} \begin{pmatrix} x_{ij(t)}x_{ji(t)} \\ x_{ij(t)}(1-x_{ji(t)}) \\ (1-x_{ij(t)})(1-x_{ji(t)}) \end{pmatrix} \left(\mathbf{1} - \frac{|z_{i(t)} - z_{j(t)}|}{max_{ij}|z_{i(t)} - z_{j(t)}|}\right)}{\sum_{i<j} x_{ij(t)}}. \quad (7)$$

*Moran's I & Geary's c*. We can also assess the degree of behavior clustering within the network by measuring network autocorrelation. This concept originated from the geospatial statistics literature, and is adopted to measure the phenomenon that the behavior of a pair of actors is typically more similar when they are connected than when they are not (Doreian 1989; Leenders 1997; Steglich, Snijders, and Pearson 2010). Two measures are generally utilized to study the magnitude of behavior similarity at the dyadic level: Moran's *I* (1948) and Geary's *c* (1954).

For Moran's *I*, the statistics is

$$A_{I(t)} = \frac{n\sum_{ij} x_{ij(t)}(z_{i(t)}-\bar{z}_{(t)})(z_{j(t)}-\bar{z}_{(t)})}{(\sum_{ij} x_{ij(t)})(\sum_{i}(z_{i(t)}-\bar{z}_{(t)})^2)}. \quad (8)$$

And for Geary's *c*, the statistic is

$$A_{c(t)} = \frac{(n-1)\sum_{ij} x_{ij(t)}(z_{i(t)}-z_{j(t)})^2}{2(\sum_{ij} x_{ij(t)})(\sum_{i}(z_{i(t)}-\bar{z}_{(t)})^2)}. \quad (9)$$

In equations (8) and (9), $n$ equals to the number of individuals in the sample, $x_{ij(t)}$ captures the presence of a tie between individuals $i$ and $j$, $z_{i(t)}$ and $z_{j(t)}$ represent the behavior values of individual $i$ and individual $j$, and $\bar{z}_{(t)}$ is denoted as the mean of behavior at time $t$.

## Method for Adequacy Checking of Behavior Dynamics

Our approach here is to use the standard Monte Carlo testing framework already widely employed for evaluating network models. The general idea of the approach is to estimate parameters for a given model, then simulate draws from the fitted model conditional on the same covariates (and, possibly, starting values) as the observed data. Indices of interest are calculated on the simulated data sets, and the quantiles of the observed index values are examined within the distribution of the simulated index values. Typically, models are regarded as adequate with respect to a given index when the observed index value is within a specified central simulation interval (e.g., 90% in Snijders 2003, and 95% in Hunter et al. 2008b). Hunter et al. (2008b) refer to this method as "graphical" GOF testing, since one can assess quality of fit with diagnostic plots (e.g., box plots for GOF testing of ERG models and violin plot for GOF testing of SAB models, see more details in the next section); visualization is not required for the procedure, however, and indeed it is a standard type of Monte Carlo test.

One common challenge with this procedure is that the number of statistics to be assessed may become quite large, motivating the use of lower-dimensional summary statistics. The

Mahalanobis distance was proposed by Wang et al. (2009) as a tool for simplifying GOF testing of ERG family models, and it was also extended to SAB modeling by Lospinoso (2012). For an observed vector of network property $A(X)$ with covariance matrix $Cov(X) = \sum$ and expected value $E(X) = \mu$, the Mahalanobis distance statistic is represented as

$$D(X) = \sqrt{(A(X) - \mu)^T \sum^{-1} (A(X) - \mu)}. \quad (10)$$

Since neither the values ($\mu$, $\sum$) nor the null distribution of $D^2(X)$ are known, the cumulative density of $D^2(X)$ is inferred from the simulated values $\hat{\mu}$ and $\hat{\sum}$. A chi-square test (Wang et al. 2009) or Monte Carlo test (Lospinoso 2012) is then performed and the *p*-value arising from the test is reported to test the null hypothesis that the specific network or behavior feature from the observed data are distributed according to that of the simulated ones. If the null hypothesis is rejected, it is a potential indicator of poor goodness of fit, and the specified model might have unreliable estimated parameters and/or need improvement in some fashion to better account for the variation in the observed data.

## Application to an Empirical Example

We apply the Mahalanobis distance testing method with the indices introduced in previous sections for model adequacy checking in the context of a typical SAB case: the Glasgow data set from *The Teenage Friends and Lifestyle Study* (Bush, West & Michell 1997, Michell and West 1996, Pearson and Michell 2000, Pearson and West 2003) which is publicly available from the Siena website.[7] 160 pupils were asked about their gender, age, pocket money per month, friendship networks, tobacco use, preferred music styles, and other elements over their second, third, and fourth year at a secondary school in Glasgow. We first estimate the SAB model for co-evolution of friendship networks and tobacco use behavior.[8] Score-type tests are utilized to aid in specifying the model. Parameters for the Model 1 are shown in Table 1.

<<<Table 1 about here>>>

In the network equation, the tobacco use similarity effect is estimated at 0.32 (s.e. = 0.16), suggesting pupils tended to select others with similar tobacco use behavior as friends. Pupils were also found to avoid forming a new tie when it would create more out-going ties. Instead, they preferred reciprocating ties and being located in transitive triplets. The interaction between reciprocity and transitivity is negative. They were also more inclined to link to someone with the same gender. In the behavior equation, the average similarity effect is estimated at 3.11 (s.e. = 1.06), suggesting that pupils tended to match their friends' tobacco use behavior. These parameters suggest several mechanisms driving the co-evolution of friendship and tobacco use behavior in this sample, but to what extent does this model actually capture the respondents' behavioral dynamics?

We begin by assessing the model adequacy of the behavior distribution at wave 3, as shown in Figure 1. This figure (and the remaining figures) uses violin plots (Hintze and Nelson 1998) to display the statistics from the 1,000 simulations based on the model. Each violin plot combines box plots (showing the median and the interquartile range) and kernel density plots, with the dashed gray lines giving the 95% simulation interval band. The observed statistics are indicated by the solid red lines on the violin plots. In this figure, the x-axis displays the three values of the behavior (tobacco use) measure on an ordinal scale. The model slightly overproduces non-smokers (value of 1 on the left side of the plot) and the observed number of non-smokers is outside of the 95% simulation interval. For occasional smokers (value of 2) and regular smokers (value of 3) the solid red line indicates that although the observed numbers in the sample are less than the medians of these simulated draws, they are nonetheless contained within the 95% confidence intervals. Thus, the model is doing a good job of reproducing

occasional and regular smokers at this time point. And the overall *p* value of 0.312 calculated from the Monte Carlo test of Mahalanobis Distance statistics indicates that the observed behavior distribution is not especially extreme compared to what would be expected from the estimated model. Thus, this model is plausibly reproducing the number of smokers at each level of the behavior.

<<<Figure 1 about here>>>

In Figure 2, we assess the adequacy of the fitted model to reproduce behavioral transitions between wave 2 and wave 3. The 9 violin plots along the x-axis indicate all the possible transitional patterns of three levels of tobacco use behavior over two time points. As can be seen, all observed data are within the 95% confidence intervals of the simulated draws for each of these transition levels. The overall *p*-value of 0.201 indicates that the observed behavioral transitions between smoking states at earlier time point and later time point are not strongly atypical of what would be expected given the fitted model.

<<<Figure 2 about here>>>

In Figure 3 we assess model adequacy for the total extent of behavior change between wave 2 and wave 3. The left-most violin plot in this figure shows the number of pupils who decreased 2 levels on the tobacco use behavior scale during the time period (going from regular smokers to non-smokers). The second violin plot shows the distribution of those decreasing 1 level on the tobacco use behavior scale (i.e., either from 3 to 2, or 2 to 1), etc. The observed number of pupils within each of these level changes is contained within the 95% confidence intervals, and the *p*-value of 0.088 also suggests compatibility of the model with the data.

<<<Figure 3 about here>>>

We next focus on how well the model captures behavior based on various network statistics. In Figure 4, we display the average out-degree based on various levels of tobacco use. Our model is accurately capturing the mean out-degree for individuals at all three levels of tobacco use. The overall *p*-value of 0.797 again indicates that the data is not strongly atypical of what the model would be expected to produce.

<<<Figure 4 about here>>>

In parallel with Figure 4, Figure 5 displays the average in-degree based on various levels of tobacco use. For each level of tobacco use behavior, the observed in-degree in the sample is contained within the 95% confidence intervals of our simulated draws from the network. The overall *p*-value of 0.521 again suggests compatibility with the model.

<<<Figure 5 about here>>>

Figure 6 assesses network-based behavior similarity in three ways. The left-most violin plot displays the goodness of fit for edgewise homophily. The sample value for this measure is contained within the interquartile range of the simulated draws from our network; this, along with the *p*-value of 0.654, suggests that our model is doing a good job reproducing this statistic. The middle violin plot shows that the value of Moran's *I* for the observed sample (capturing the extent of clustering in the network based on tobacco use behavior) is also contained within the interquartile range of these simulated draws of the network from our model. The *p*-value of 0.63 indicates that our model is doing a good job of reproducing this characteristic of the network. The right-most violin plot shows that the value of Geary's *c* in this sample (another measure of clustering in the network based on tobacco behavior) is contained within the interquartile range of the simulated draws of the network from our model, and the *p*-value of 0.627 indicates a good fit of the model to the data.

<<<Figure 6 about here>>>

Finally, we demonstrate that our suggested GOF indices for behavior dynamics can be used to compare model fit across different model specifications by estimating a more parsimonious model and comparing it to our model. This alternative model excludes the gender related effects from both the network and behavior equations. The parameter estimates for this Model 1a are displayed in Table 1. We notice that both peer selection and peer influence have larger estimated parameters but smaller standard errors in the alternative model.

As shown in Figures S1 to S6 in the supplemental documents, the alternative model also successfully reproduces the macro/global properties of behavior distribution, behavior transition, average out-degree and in-degree by each behavior level, edgewise homophily, Moran's I, and Geary's c. However, it exhibits poor fit in the behavior change values. Thus, a model may capture some of these dimensions accurately, but it is necessary to assess model fit along a number of dimensions to better assess the model.

## Extension to a Two-mode Network

As Snijders, Lomi, and Torló (2013) pointed out, the basic models for joint (one-model) network/behavior evolution and joint one-mode network/two-mode network evolution are mathematically "quite similar." We can conceptualize activity/group/event affiliation as a collective set of multiple behavior which can take only two values – 0 (no) and 1 (yes). Therefore, it is straightforward for the desiderata and indices for model adequacy checking of behavior dynamics to be extended from one-mode network/behavior models to joint two-mode network evolution.

Using the Glasgow data set we estimate a SAB model for co-evolution of friendship networks and 16 music style preference. This model asks whether sharing similar music interest

drives friendship choice, or whether the musical preferences of one's friends drive the change in a pupil's musical preferences. The results for the network equation in Table 2 are similar to that in Table 1 except that there was no evidence in this model that pupils make friendship decisions based on homophily of music preference, as the friendship from music parameter (.03, $p > .05$) was non-significant. That is, friendship decisions are not significantly driven by sharing common preference in music styles.

<<<Table 2 about here>>>

In the behavior equation, a pair of pupils who had one music style in common were likely to obtain more styles in common (i.e., the positive 4-cycle effect). Music styles having many fans would get even more fans (i.e., the positive in-degree popularity effect), and a pupil preferring many music styles would further augment this tendency (i.e., the positive out-degree activity effect). However, some pupils had a tendency not to prefer any music style (i.e., the negative effect for out-degree at least 1), and a pupil preferring many music styles was inclined to become or remain a fan of styles not as popular (i.e., the negative out-in degree assortativity effect). Moreover, a pupil preferring a type of music style from a certain category (e.g., rock, heavy, indie, and grunge in the rock category and dance, techno, and rave in the techno category) was more likely to prefer other music styles from the same category (i.e., the positive choosing music style in the same category effect). Finally, while a pupil tended to become or remain a fan of a music style if his or her friends were fans also (i.e., the positive friendship to agreement effect), he or she was less likely to be a fan of a music style if another pupil sharing a common friend with him or her was a fan of that music style (i.e., the negative shared friendship to agreement effect).

Regarding distribution of music style preference, Figure 7 shows that the observed numbers of fans in the 16 music styles are contained within the 95% confidence intervals. The *p*-value of 0.103 also indicates the compatibility of the model to the data.

<<<Figure 7 about here>>>

We can also test the model adequacy of transition and value changes in music style preference. However, given the long list of music styles, we do not pursue this here. However, even in an instance with so many musical styles in the model, a researcher might wish to test transitions for specific musical styles of particular theoretical interest.

Figure 8 and Figure 9 show the average out-degree and the average in-degree by each music style. The observed numbers are all contained within the 95% confidence intervals. The *p*-values of 0.457 and 0.198 indicate a good model fit.

<<<Figure 8 about here>>>

<<<Figure 9 about here>>>

In two-mode networks, edgewise homophily statistics equate to the proportion of edges in which a pair of pupils shared a music style in common (i.e., $z_i = z_j = 0$, or $z_i = z_j = 1$). As shown in Figure 10, the model does a good job in simulating edges two pupils shared in common across the 16 music styles. The *p*-value is 0.122, again demonstrating a good fit.

<<<Figure 10 about here>>>

Finally, Figures 11 and 12 show that the model successfully reproduces the preference clustering in the 16 music styles measured in Moran's *I* and Geary's *c* across friendship edges. The *p*-values are greater than .05, again suggesting an satisfactory fit.

<<<Figure 11 about here>>>

<<<Figure 12 about here>>>

Poor fit implies the potential for model improvement. The GOF testing indices for both network and *behavior* dynamics over different nested or even competing models can be used to decide among different model specifications. If the model prediction departs significantly from the observed data on the GOF criteria, a better model may be preferred. Alternately, if the statistics on which the departure is observed are not critical for the model's intended purpose, then it may be reasonable to retain the current model for other reasons (e.g., parsimony, or superior performance on a dimension of primary interest). As always, such decisions must reflect a clear sense of what the model is intended to capture, and the function it is intended to serve. These may involve both practical and theoretical considerations (e.g., limits to computing power or data availability, on the one hand, or the desire to have a model that implements a specific set of hypothetical mechanisms on the other), and the same model that is considered adequate for one purpose may not be adequate for another. Clarity regarding modeling objectives is hence essential for successful model assessment.

## Discussion

In this paper we have described the importance of assessing model adequacy for dynamic models of both networks and behavior, with the Glasgow data set using SAB models. This study contributes to the current literature on model assessment – which focuses on the reproduction of mostly structural properties – by introducing basic measures of behavioral and joint behavioral/network dynamics that can be used in the adequacy checking process.

We adopted the common strategy of assessing the fit of the model by comparing the global values of the distribution of behavior in the actual observed networks to those from networks simulated based on the model parameter values and proposed four reasonable desiderata to be considered when assessing how well the model captures behavior change at the

global/macro level over time. First, the model should be able to reproduce the distribution of the behavior in the observed sample at each time point. Second, the model should be able to accurately capture *transitions* over time in the level of behavior. Third, the model should be able to capture the extent to which the behavior patterns align with key statistics of the social network. Fourth, the model should be able to capture the degree of behavior clustering among ties within the network. We proposed indices capturing each of these goals and demonstrated their use on an example dataset. We also extended these desiderata and indices to two-mode network dynamics with the same dataset. We emphasize that a model might be adequate with respect to some desiderata, but not others. Where the features not reproduced are not of substantive interest, where a model is attractive in other respects (e.g., by being highly parsimonious), or where the poorly reproduced features correspond to questionable aspects of the original data, this may not be a concern. On the other hand, failure to reproduce key features of substantive interest may suggest that the model is not capturing the desired generative process. In this case, model improvement is often warranted. It must be borne in mind that no simple, low-dimensional model of a complex, high-dimensional system is likely to reproduce the latter in all respects. Similarly, when evaluating many dimensions of model performance, some will by chance alone deviate from observation. Thus, the use of scientific judgment to weight the nature, number, and magnitude of deviations when deciding whether a given model is sufficient for a particular purpose is an inescapable part of the evaluation process.

These proposed indices are reasonably simple and easy statistics to compute. For researchers using the SAB framework with the RSiena software package, it is particularly straightforward to utilize these measures. Key scripts are provided in supplemental document 2. We have emphasized that it is important to assess how well the model is reproducing not just the

network, but also the behavior of the individuals in dynamic network and behavior models. Of course, accurately reproducing the observed data does not assure that the model is "correct". But failing to reproduce the observed data raises the concern that the model is not a good reflection of the processes that generated the data in question, and that inferences drawn from it may be suspect.

## Notes

1. Note that we are not here focused on notions of goodness of fit related to minimization of a global objective function such as the deviance or squared error, nor to questions of model selection per se. A large theoretical literature exists on these and related issues from a range of perspectives; see e.g. Wald (1950), Scholkopf and Smola (2001), Bernardo and Smith (2007).
2. Though the difficulty of specifying effective data degrees of freedom required for the AICc and BIC and questions regarding the applicability of the asymptotic foundations of the AIC, AICc, and BIC to conventional network models have limited the use of these model selection criteria.
3. Associations between behavior level and other aspects of structural position (e.g., embeddedness in reciprocated relations) can also be targets of investigation. Since there is an unlimited number of such interactions, and their use is theory-specific, we do not consider them in detail here. However, we note that, when screening for potential effects, the large number of possible structure/behavior interactions makes explicit estimation of all possible models prohibitive. To this end, *score tests* are often recommended to identify potentially significant effects (relative to a base model) in a computationally practical way. More details of score-type tests are available in Ripley et al. (2019), and a theoretical exposition is available in Schweinberger (2012).

4. These indices are implemented with the ergm software package (Hunter et al. 2008a).

5. These indices are implemented with the RSiena software packages (Ripley et al. 2019).

6. This index, however, is not yet implemented in the RSiena software package as of this writing.

7. More detailed description of the Glasgow data set is available at http://www.stats.ox.ac.uk/~snijders/siena/Glasgow_data.htm. The data set can be downloaded from http://www.stats.ox.ac.uk/~snijders/siena/Glasgow_data.zip.

8. We adopt both *t* statistics for deviations from targets (i.e., ideally less than 0.10 for each parameter) and the overall maximum convergence ratio (i.e., ideally less than 0.25) to satisfy model convergence.

## References


Akaike, H. 1973. "Information Theory as an Extension of the Maximum Likelihood Principle." Pp. 267-81 in *Second International Symposium on Information Theory*, edited by B. N. Petrov and F. Csaki. Budapest: Akademiai Kiado.

Almquist, Z. W., and C. T. Butts. 2013. "Dynamic Network Logistic Regression: A Logistic Choice Analysis of Inter- and Intra-Group Blog Citation Dynamics in the 2004 US Presidential Election." *Political Analysis* 21: 430-48.

Almquist, Z. W., and C. T. Butts. 2014. "Logistic Network Regression for Scalable Analysis of Networks with Joint Edge/Vertex Dynamics." *Sociological Methodology* 44: 273-321.

Bernardo, J. M., and A. F. M. Smith. 2007. *Bayesian Theory*. New York, NY: Wiley.

Box, G. E. P., and N. R. Draper. 1987. *Empirical Model Building and Response Surfaces*. New York, NY: John Wiley & Sons.

Bush, H., P. West, and L. Michell. 1997. "The Role of Friendship Groups in the Uptake and Maintenance of Smoking amongst Pre-Adolescent and Adolescent Children: Distribution of Frequencies." Working Paper No. 62. MRC Medical Sociology Unit Glasgow.

Carley, K. 1991. "A Theory of Group Stability." *American Sociological Review* 56: 331-54.

Christakis, N. A., and J. H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *New England Journal of Medicine* 357: 370-79.

Coleman, J. S. 1964. *Introduction to Mathematical Sociology*. New York, NY: Free Press of Glencoe/Collier-Macmillan.

Coleman, J. S. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University.

Coleman, J. S., H. Katz, and H. Menzel. 1966. *Medical Innovation: A Diffusion Study*. Indianapolis, IN: Bobbs-Merill.

Dishion. T. J., and L. D. Owen. 2002. "A Longitudinal Analysis of Friendships and Substance Use: Bidirectional Influence from Adolescence to Adulthood." *Developmental Psychology* 38: 480-91.

Doreian, P. 1989. “Network Autocorrelation Models: Problems and Prospects.” Pp. 369-389 in *Spatial Statistics: Past, Present, Future*, edited by D. A. Griffith. Ann Arbor, MI: Michigan Document Services.

Fink, B., and K.-P. Wild. 1995. “Similarities in Leisure Interests: Effects of Selection and Socialization in Friendships.” *The Journal of Social Psychology* 135: 471-82.

Geary, R. C. 1954. “The Contiguity Ratio and Statistical Mapping.” *Incorporated Statistician* 5: 115-45.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2003. *Bayesian Data Analysis, 2nd Edition*. Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., and X. L. Meng. 1996. “Model Checking and Model Improvement.” Pp. 189-201 in *Practical Markov Chain Monte Carlo*, edited by W. Gilks, S. Richardson, and D. Spiegelhalter. London: Chapman & Hall.

Gelman, A., X. L. Meng, and H. Stern. 1996. “Posterior Predictive Assessment of Model Fitness via Realized Discrepancies.” *Statistica Sinica* 6: 733-807

Hanneke, S., W. Fu, and E. P. Xing. 2010. “Discrete Temporal Models of Social Networks.” *Electronic Journal of Statistics* 4: 585-605.

Harris, K. M., C. T. Halpern, E. Whitsel, J. Hussey, J. Tabor, P. Entzel, and J. R. Udry. 2009. *The National Longitudinal Study of Adolescent Health: Research Design*. Available online at http://www.cpc.unc.edu/projects/addhealth/design.

Hintze, J., and R. Nelson. 1998. “Violin Plots: A Box Plot-Density Trace Synergism.” *American Statistician* 52: 181-4.

Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris. 2008a. “ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks.” *Journal of Statistical Software* 24: 1-29.

Hunter, D. R., S. M. Goodreau, and M. S. Handcock. 2008b. “Goodness of Fit of Social Network Models.” *Journal of the American Statistical Association* 103: 248-58.

Kandel, D. B. 1978. “Homophily, Selection and Socialization in Adolescent Friendships.” *American Journal of Sociology* 84: 427-36.

Kandel, D. B. 1985. “On Processes of Peer Influences in Adolescent Drug Use: A Developmental Perspective.” *Advances in Alcohol & Substance Abuse* 4: 139-63.

Krivitsky, P. N., and M. S. Handcock. 2014. “A Separable Model for Dynamic Networks.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76: 29-46.

Leenders, R. T. A. J. 1997. “Longitudinal Behavior of Network Structure and Actor Attributes: Modeling Interdependence of Contagion and Selection.” Pp. 165-184 in *Evolution of Social Networks*, edited by P. Doreian, and F. N. Stokman. New York, NY: Gordon & Breach.

Lospinoso, J. A. 2012. *Statistical Models for Social Network Dynamics*. PhD thesis, University of Oxford, U.K.

Lospinoso, J. A., and T. A. B. Snijders. 2011. “Goodness of Fit for Social Network Dynamics.” Presentation at the Sunbelt XXXI on February 12, 2011, St. Pete's Beach, Florida.

Macy, M. W., J. A. Kitts, A. Flache, and S. Benard. (2003). “Polarization in Dynamic Networks: A Hopfield Model of Emergent Structure.” Pp. 162-73 in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, edited by R. Breiger, K. Carley, and P. Pattison. Washington, DC: National Academies Press.

Mark, N. 1998. “Beyond Individual Differences: Social Differentiation from First Principles.” *American Sociological Review* 63: 309-30.

Michell, L., and P. West. 1996. “Peer Pressure to Smoke: The Meaning Depends on the Method.” *Health Education Research* 11: 39-49.

Mislove, A., M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. 2007. “Measurement and Analysis of Online Social Networks.” Pp. 29-42 in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*. San Diego, CA: ACM.

Moran, P. A. P. 1948. “The Interpretation of Statistical Maps.” *Journal of the Royal Statistical Society, Series B* 10: 245-51.

Osgood, D. W., D. T. Ragan, L. Wallace, S. D. Gest, M. E. Feinberg, and J. Moody. 2013. “Peers and the Emergence of Alcohol Use: Influence and Selection Processes in Adolescent Friendship Networks.” *Journal of Research on Adolescence* 23: 500-12.

Pearson, M., and L. Michell. 2000. “Smoke Rings: Social Network Analysis of Friendship Groups, Smoking and Drug-Taking.” *Drugs: Education, Prevention and Policy* 7: 21-37.

Pearson, M., and P. West. 2003. “Drifting Smoke Rings: Social Network Analysis and Markov Processes in a Longitudinal Study of Friendship Groups and Risk-Taking.” *Connections* 25: 59-76.

Purta, R., S. Mattingly, L. Song, O. Lizardo, D. S. Hachen, C. Poellabauer, and A. Striegel. 2016. “Experiences Measuring Sleep and Physical Activity Patterns Across A Large College Cohort with Fitbits.” Pp. 28-35 in *Proceedings of the 2016 ACM International Symposium on Wearable Computers*. New York, NY: ACM.

Ripley, R. M., T. A. B. Snijders, Z. Boda, A. Vörös, and P. Preciado. 2019. *Manual for Siena Version 4.0 (Version April 9, 2019)*. University of Oxford, Department of Statistics, Nueld College. Available online at http://www.stats.ox.ac.uk/~snijders/siena/.

Robins, G., and P. Pattison. 2001. “Random Graph Models for Temporal Processes in Social Networks.” *Journal of Mathematical Sociology* 25: 5-41.

Robins, G., P. Pattison, and P. Elliott. 2001. “Network Models for Social Influence Processes.” *Psychometrika* 66: 161-89.

Scholkopf, B., and A. J. Smola. 2001. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press.

Schwarz, G. E. 1978. “Estimating the Dimension of a Model.” *Annals of Statistics* 6: 461-64.

Schweinberger, M. 2012. “Statistical Modelling of Network Panel Data: Goodness of Fit.” *British Journal of Mathematical and Statistical Psychology* 65: 263-81.

Snijders, T. A. B. 1996. “Stochastic Actor-Oriented Models for Network Change.” *Journal of Mathematical Sociology* 21: 149-72.

Snijders, T. A. B. 2003. “Accounting for Degree Distributionsi Empirical Analysis of Network Dynamics.” Pp. 146-61 in *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, edited by R. Breiger, K. Carley, and P. Pattison. Washington, DC: National Academies Press.

Snijders, T. A. B. 2011. “Network Dynamics.” Pp. 501-13 in *The SAGE Handbook of Social Network Analysis*, edited by J. Scott, and P. J. Carrington. Thousand Oaks, CA: Sage.

Snijders, T. A. B., A. Lomi, and V. J. Torló. (2013). “A Model for the Multiplex Dynamics of Two-Mode and One-Mode Networks, with an Application to Employment Preference, Friendship, and Advice.” *Social Networks* 35: 265-276.

Snijders, T. A. B., and C. E. G. Steglich. 2013. “Representing Micro-Macro Linkages by Actor-Based Dynamic Network Models.” *Sociological Methods & Research* 2013. doi: 10.1177/0049124113494573

Snijders, T. A. B., C. E. G. Steglich, and M. Schweinberger. 2007. "Modeling the Co-Evolution of Networks and Behavior." Pp. 41-71 in *Longitudinal Models in the Behavioral and Related Sciences*, edited by: K. van Montfort, J. Oud, and A. Satorra. Mahwah, NJ: Lawrence Erlbaum.

Snijders, T. A. B., G. G. van de Bunt, and C. E. G. Steglich. 2010. "Introduction to Stochastic Actor-Based Models for Network Dynamics." *Social Networks* 32: 44-60.

Snijders, T. A. B., J. Koskinen, and M. Schweinberger. 2010. "Maximum Likelihood Estimation for Social Network Dynamics." *Annals of Applied Statistics* 4: 567-88.

Steglich, C. E. G., P. Sinclair, J. Holliday, and L. Moore. 2012. "Actor-Based Analysis of Peer Influence in A Stop Smoking In Schools Trial (ASSIST)." *Social Networks* 34: 359-69.

Steglich, C. E. G., T. A. B. Snijders, and M. Pearson. 2010. "Dynamic Networks and Behavior: Separating Selection from Influence." *Sociological Methodology* 40: 329-93.

Striegel, A., S. Liu, L. Meng, C. Poellabauer, D. Hachen, and O. Lizardo. 2013. "Lessons Learned from the Netense Smartphone Study." Pp. 51-56 in *HotPlanet '13 Proceedings of the 5th ACM Workshop*. New York, NY: ACM.

Wald, A. 1950. *Statistical Decision Functions*. New York, NY: John Wiley and Sons.

Wang, C., C. T. Butts, J. R. Hipp, R. Jose, and C. M. Lakon. 2016. "Multiple Imputation for Missing Edge Data: A Predictive Evaluation Method with Application to Add Health." *Social Networks* 45: 89-98.

Wang, P., K. Sharpe, G. L. Robins, and P. E. Pattison. 2009. "Exponential Random Graph (P*) Models for Affiliation Networks." *Social Networks* 31: 12-25.

Wasserman, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.

West, P., and H. Sweeting. 1995. *Background Rationale and Design of the West of Scotland 11-16 Study*. Working Paper No. 52. MRC Medical Sociology Unit Glasgow, Glasgow.

**Tables and Figures**

**Table 1.** SAB Model for Joint Friendship Networks and Tobacco Use Behavior Evolution Using the Glasgow Data Set

| Effect name | Model 1 | | Model 1a | |
|---|---|---|---|---|
| | beta | s.e. | beta | s.e. |
| Network decision: selection processes | | | | |
| Constant friendship rate (period 1) | 14.25*** | 1.06 | 14.43*** | 1.37 |
| Constant friendship rate (period 2) | 11.88*** | 1.23 | 11.82*** | 1.16 |
| Out-degree (density) | -2.85*** | 0.07 | -2.71*** | 0.05 |
| Reciprocity | 2.42*** | 0.13 | 2.57*** | 0.12 |
| Transitive triplets | 0.79*** | 0.05 | 0.83*** | 0.04 |
| Transitive reciprocated triplets | -0.57*** | 0.11 | -0.63*** | 0.08 |
| 3-cycles | 0.12 | 0.11 | 0.17 | 0.10 |
| Gender alter (female = 1) | -0.12 | 0.11 | | |
| Gender ego (female = 1) | 0.05 | 0.12 | | |
| Gender similarity (female = 1) | 0.71*** | 0.09 | | |
| Tobacco use alter | 0.08 | 0.07 | 0.08 | 0.09 |
| Tobacco use ego | -0.04 | 0.08 | -0.04 | 0.08 |
| Tobacco use similarity | 0.32* | 0.16 | 0.41** | 0.15 |
| Behavior decision: influence processes | | | | |
| Rate tobacco use (period 1) | 4.42*** | 1.08 | 4.31* | 2.10 |
| Rate tobacco use (period 2) | 3.80* | 1.51 | 3.71* | 1.80 |
| Tobacco use linear shape | -3.28*** | 0.86 | -3.30*** | 0.77 |
| Tobacco use quadratic shape | 2.73*** | 0.41 | 2.72*** | 0.37 |
| Tobacco use average similarity | 3.11** | 1.06 | 3.19** | 0.98 |
| Tobacco use in-degree | 0.06 | 0.15 | 0.04 | 0.12 |
| Tobacco use out-degree | 0.11 | 0.25 | 0.15 | 0.22 |
| Tobacco use: effect from gender (female = 1) | -0.10 | 0.28 | | |
| Tobacco use: effect from parent smoking | -0.20 | 0.29 | -0.20 | 0.27 |
| Tobacco use: effect from sibling smoking | 0.18 | 0.40 | 0.18 | 0.41 |
| Tobacco use: effect from pocket money per month | 0.01 | 0.02 | 0.01 | 0.02 |

* Two-sided p<0.05; ** Two-sided p<0.01; *** Two-sided p<0.001; $n$ = 160

**Table 2.** SAB Model for Joint Friendship Networks and Music Style Preference Evolution Using the Glasgow Data Set

| Effect name | beta | s.e. |
|---|---|---|
| Network decision: selection processes | | |
| Constant friendship rate (period 1) | 11.70*** | 0.94 |
| Constant friendship rate (period 2) | 10.81*** | 1.07 |
| Friendship: out-degree (density) | -3.05*** | 0.08 |
| Friendship: reciprocity | 2.49*** | 0.11 |
| Friendship: transitive triplets | 0.66*** | 0.04 |
| Friendship: transitive reciprocated triplets | -0.50*** | 0.09 |
| Friendship: 3-cycles | 0.01 | 0.09 |
| Friendship: gender alter (female = 1) | -0.13 | 0.09 |
| Friendship: gender ego (female = 1) | 0.08 | 0.10 |
| Friendship: gender similarity (female = 1) | 0.85*** | 0.09 |
| Friendship: from music agreement | 0.03 | 0.03 |
| Behavior decision: influence processes | | |
| Constant music rate (period 1) | 5.69*** | 0.48 |
| Constant music rate (period 2) | 5.55*** | 0.42 |
| Music: out-degree (density) | -1.89*** | 0.15 |
| Music: 4-cycles | 0.02*** | 0.00 |
| Music: in-degree - popularity | 0.02*** | 0.00 |
| Music: out-degree - activity | 0.18*** | 0.03 |
| Music: out-degree >=1 | -1.93*** | 0.36 |
| Music: out-in degree assortativity | -0.01*** | 0.00 |
| Music: gender ego (female = 1) | 0.03 | 0.09 |
| Music: gender ego-in-alter distance 2 similarity (female = 1) | 0.42 | 0.43 |
| Music: 4-cycles same gender (female = 1) | 0.00 | 0.01 |
| Music: age ego | -0.16 | 0.11 |
| Music: pocket money per month ego | 0.00 | 0.01 |
| Music: in-degree friendship activity | 0.03 | 0.04 |
| Music: out-degree friendship activity | -0.06 | 0.06 |
| Music: Choosing music style in the same category | 0.14*** | 0.03 |
| Music: friendship to agreement | 0.69*** | 0.11 |
| Music: shared friendship (1) to agreement | -0.08** | 0.02 |

** Two-sided p<0.01; *** Two-sided p<0.001; $n$ = 160

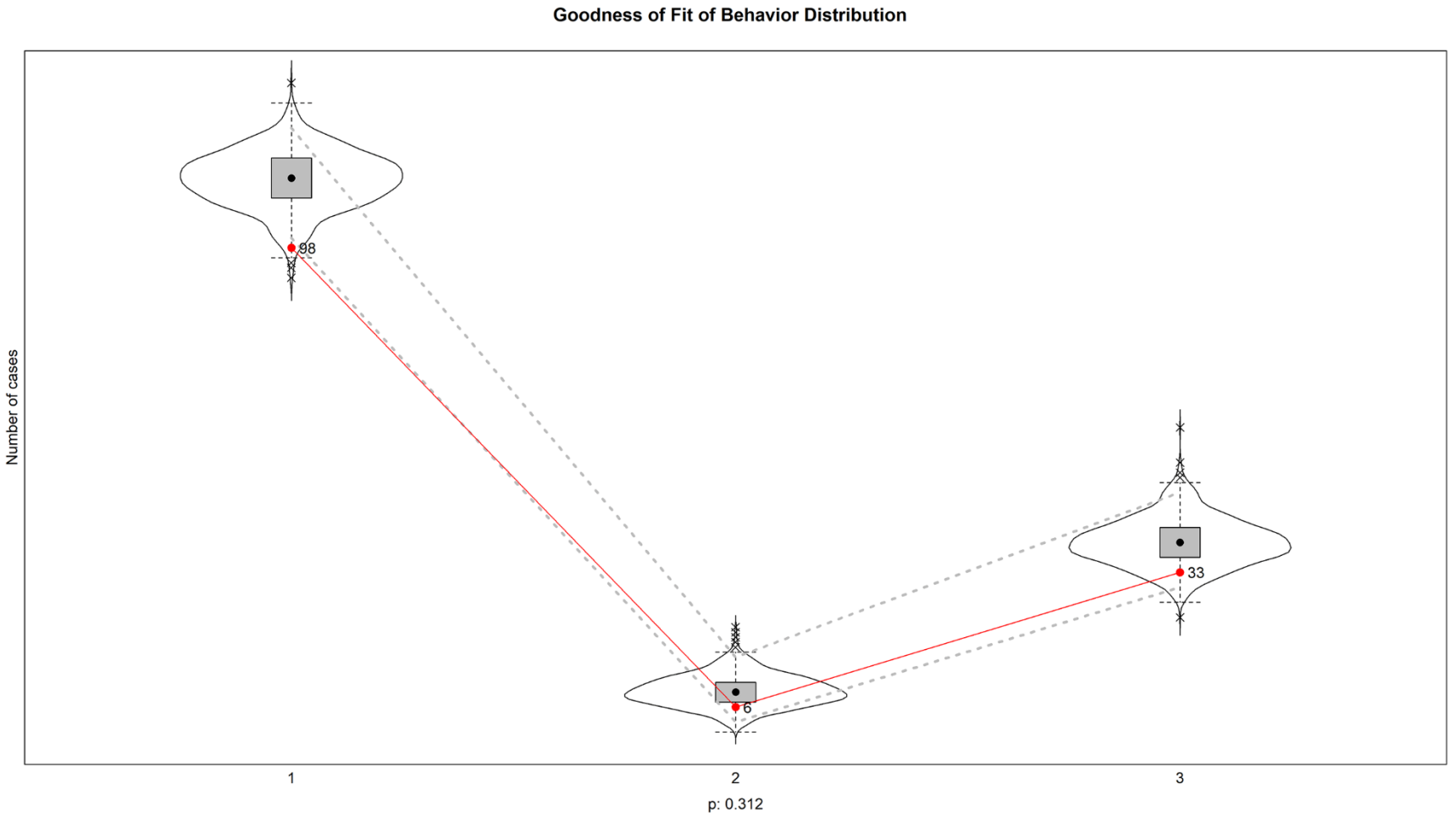


**Figure 1.** GOF testing for tobacco use behavior distribution.

*Note:* 1 - non; 2 - occasional; 3 - regular, i.e. more than once per week

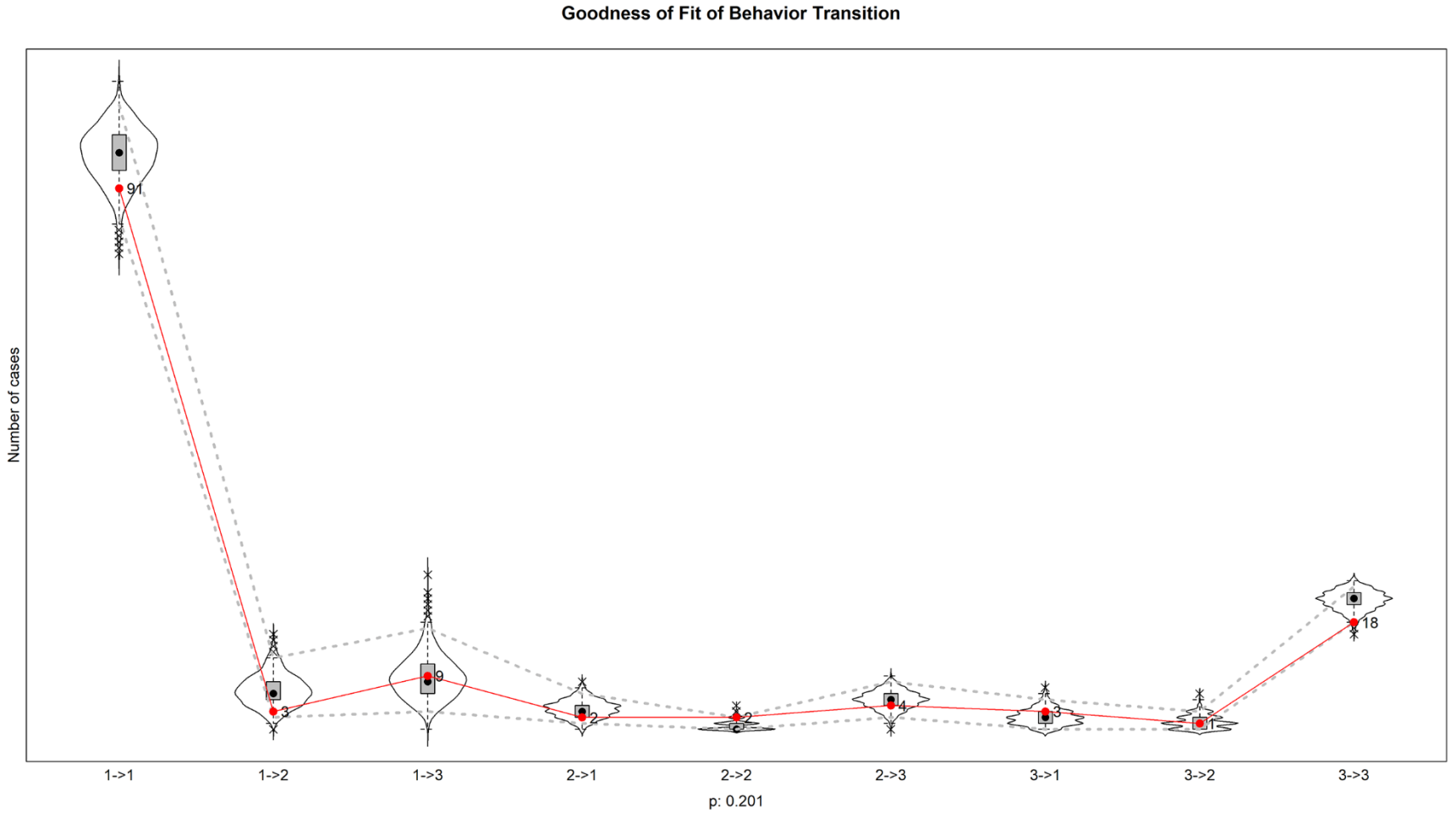


**Figure 2.** GOF testing for tobacco use behavior transition.

*Note:* 1 - non; 2 - occasional; 3 - regular, i.e. more than once per week

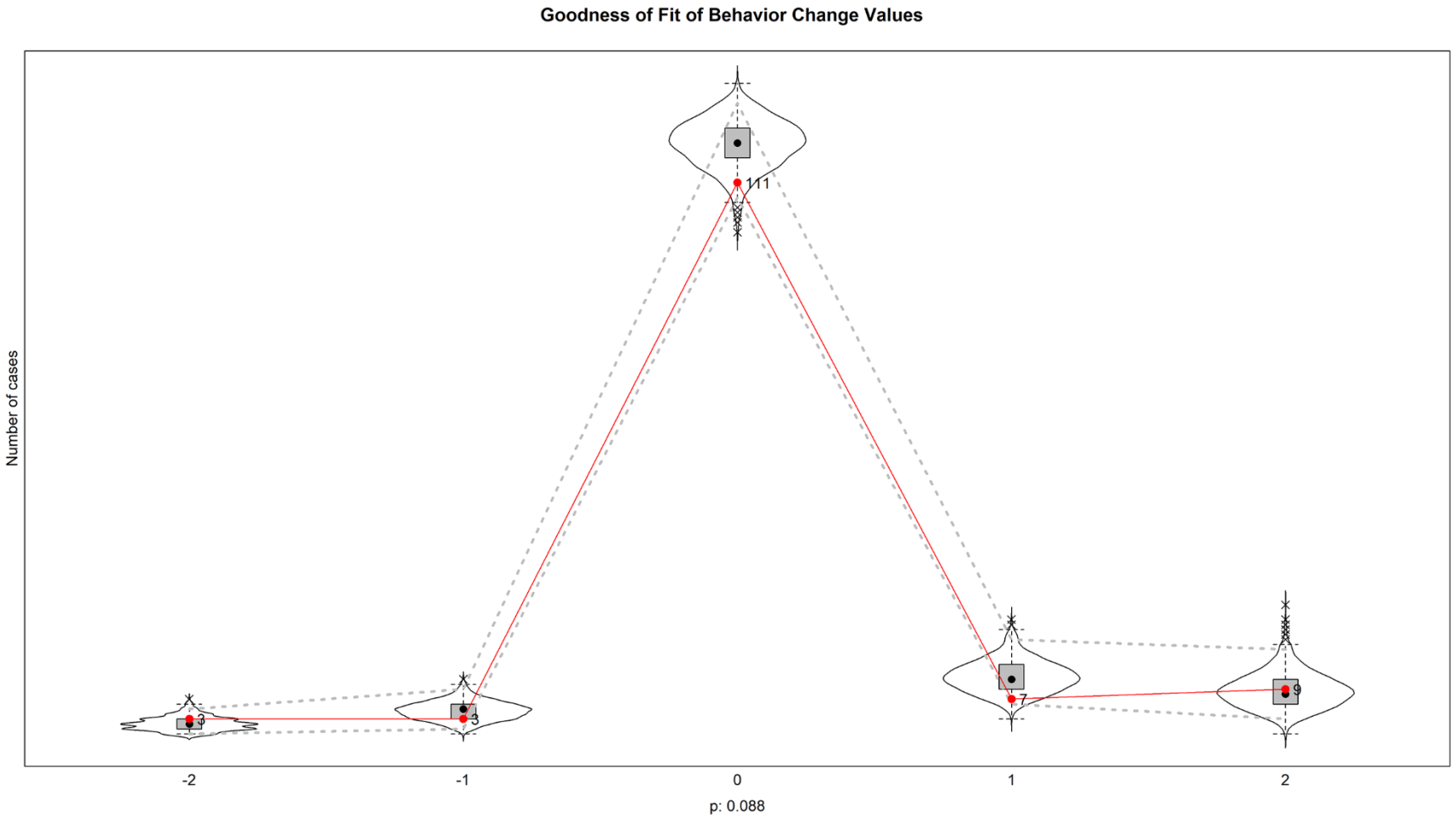


**Figure 3.** GOF testing for tobacco use behavior change statistics.

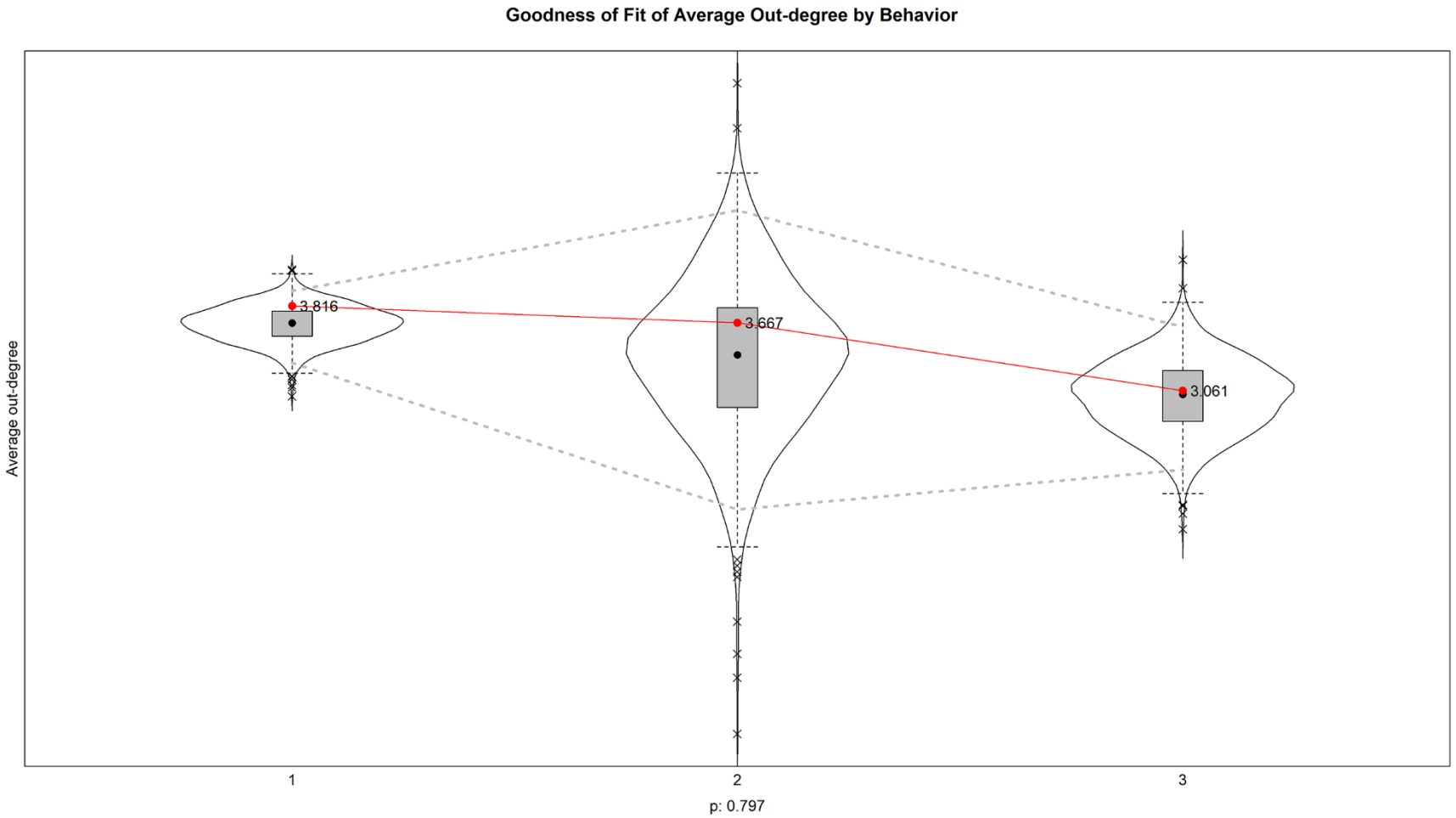


**Figure 4**. GOF testing for behavior related average out-degree.

*Note:* 1 - non; 2 - occasional; 3 - regular, i.e. more than once per week

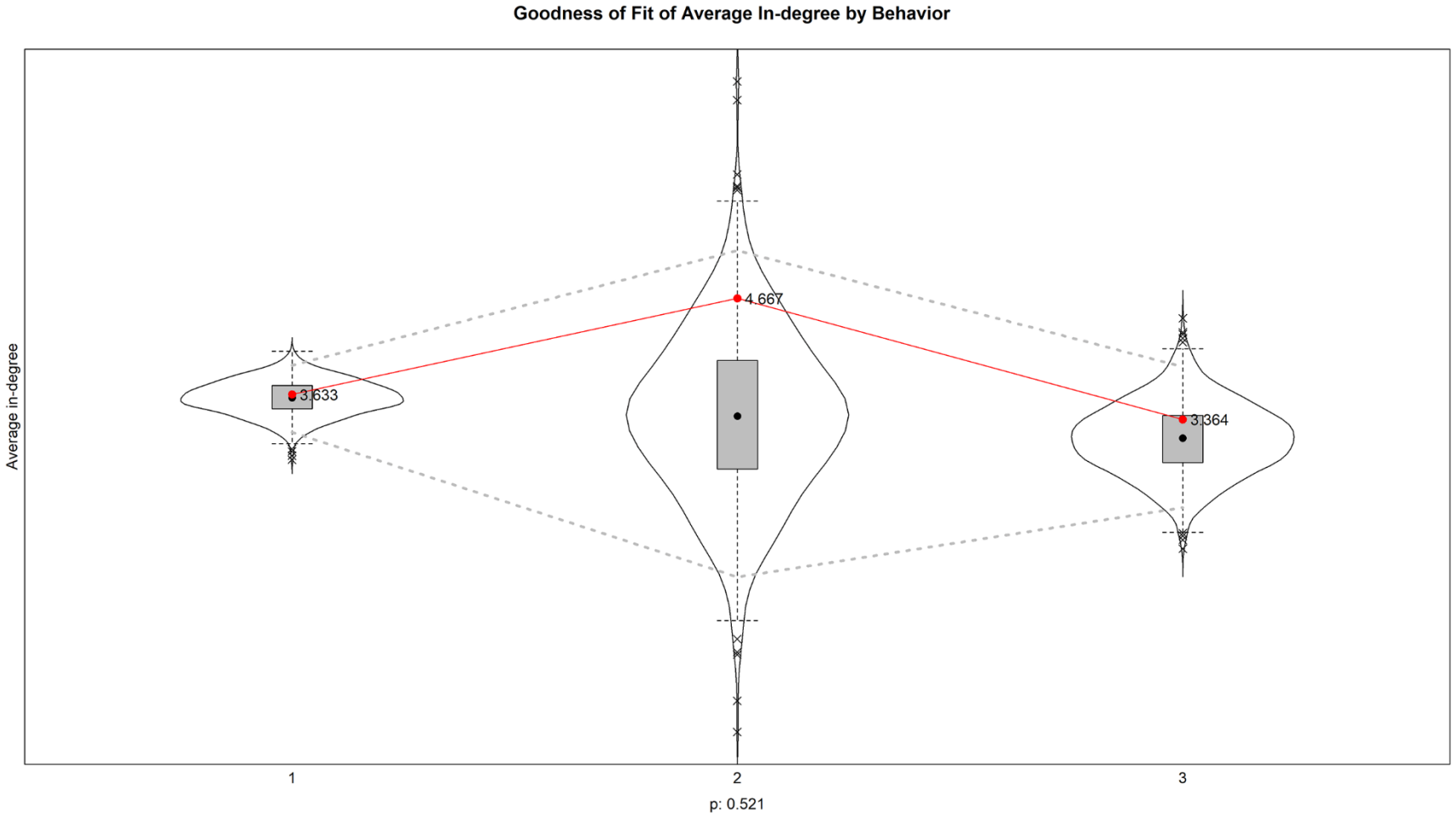


**Figure 5.** GOF testing for behavior related average in-degree.

*Note:* 1 - non; 2 - occasional; 3 - regular, i.e. more than once per week

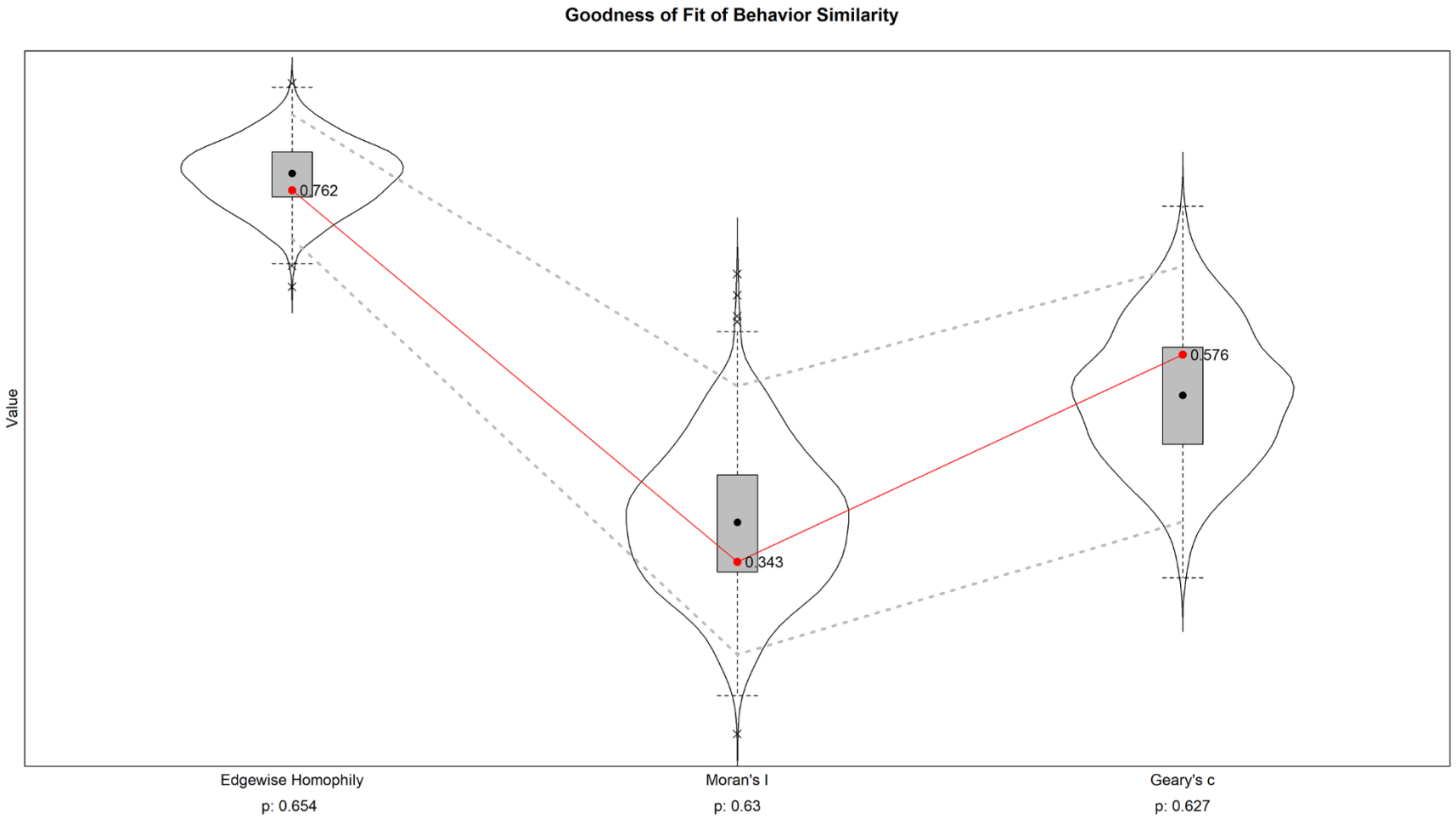


**Figure 6.** GOF testing for behavior similarity in tobacco use.

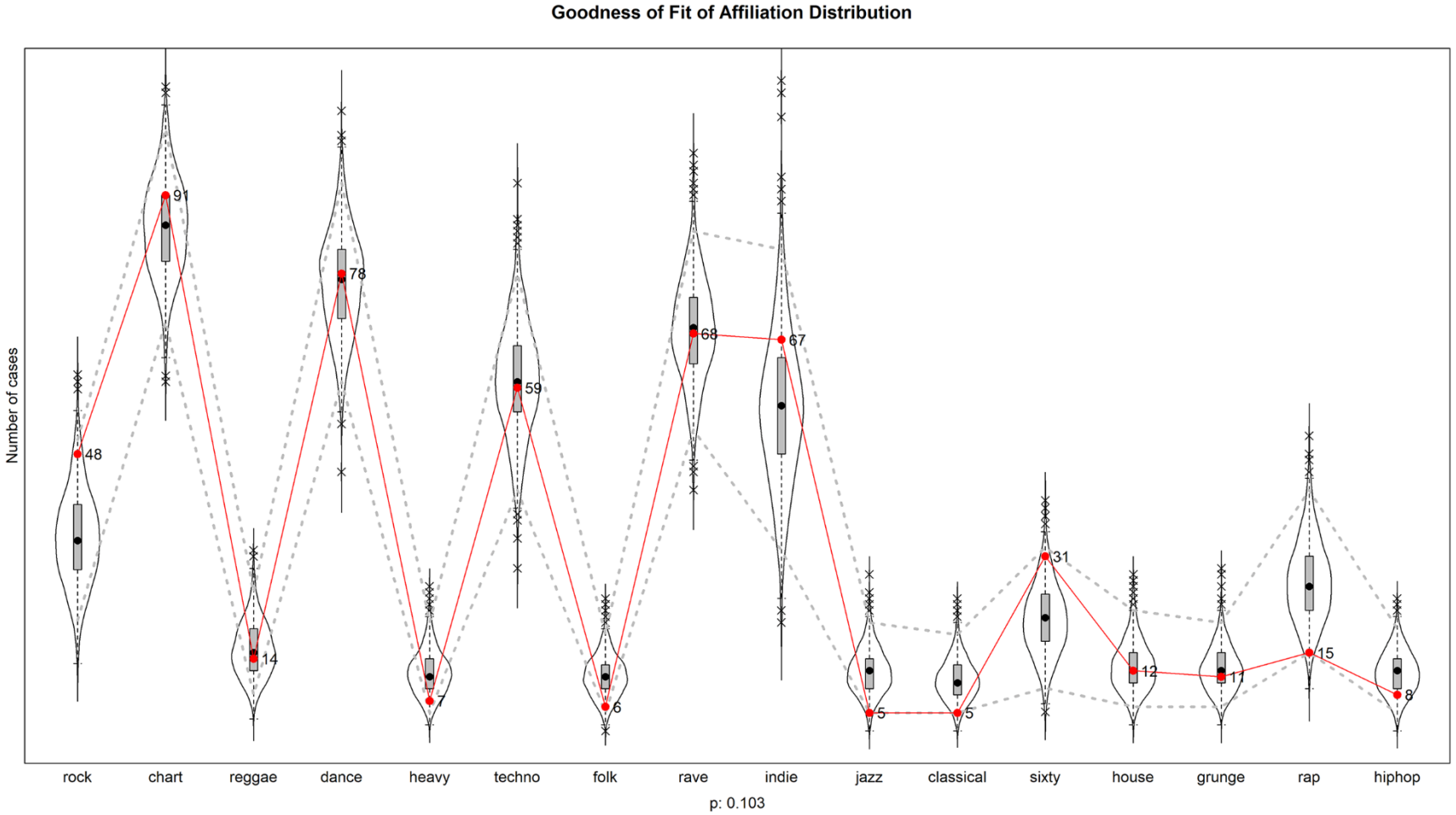


**Figure 7.** GOF testing for distribution of music style preference.

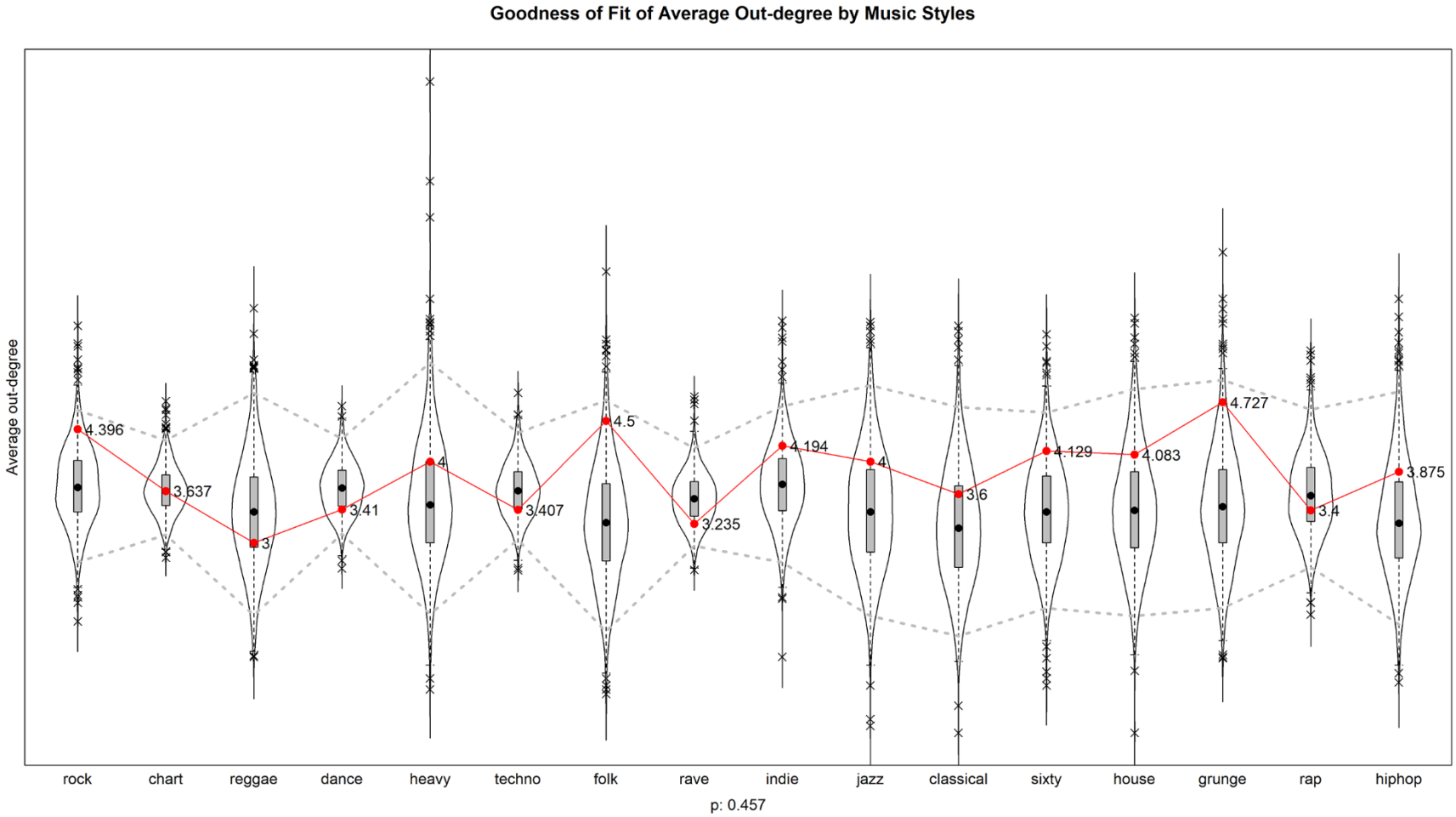


**Figure 8.** GOF testing for average out-degree by music style preference.

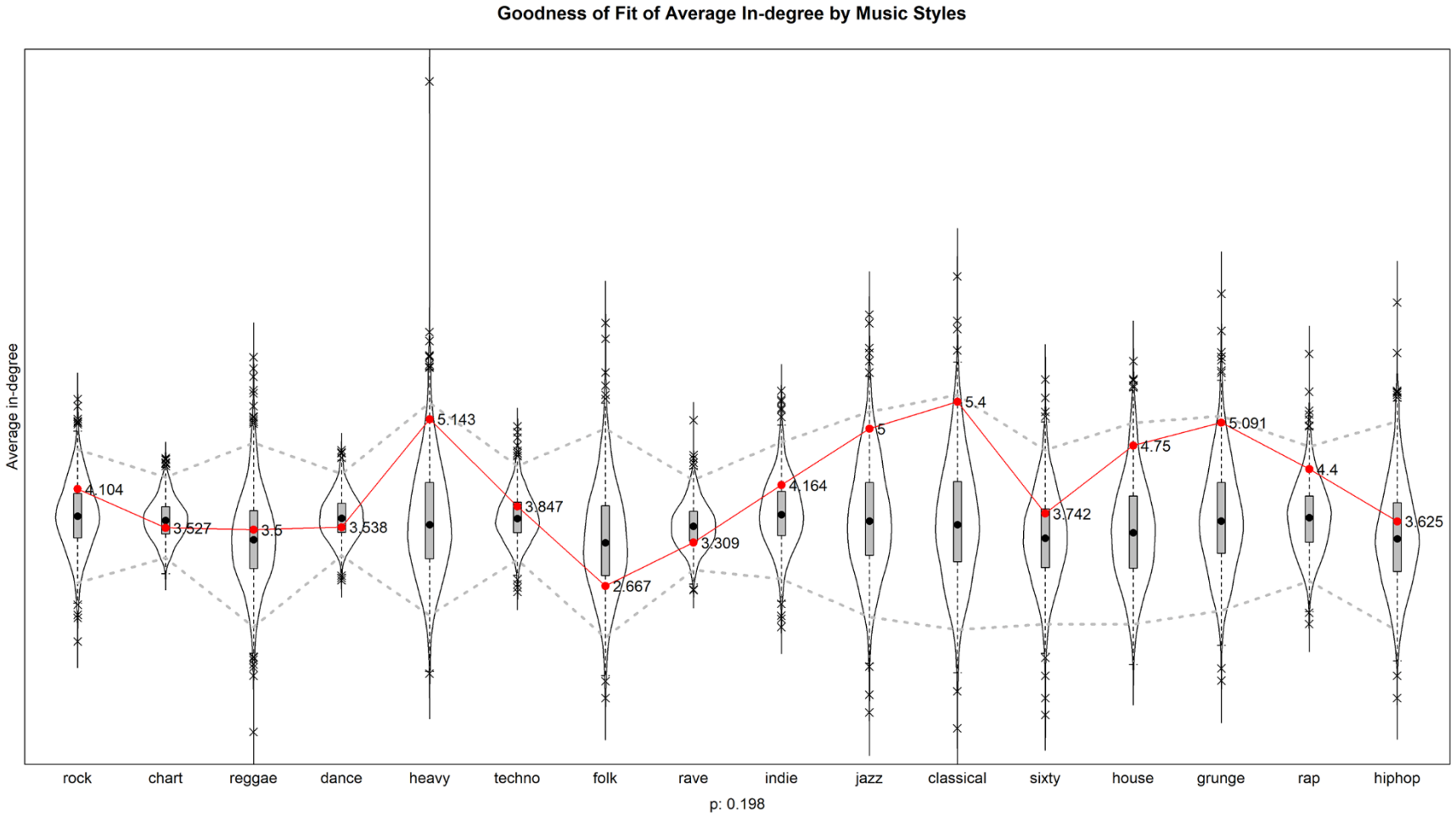


**Figure 9.** GOF testing for average in-degree by music style preference.

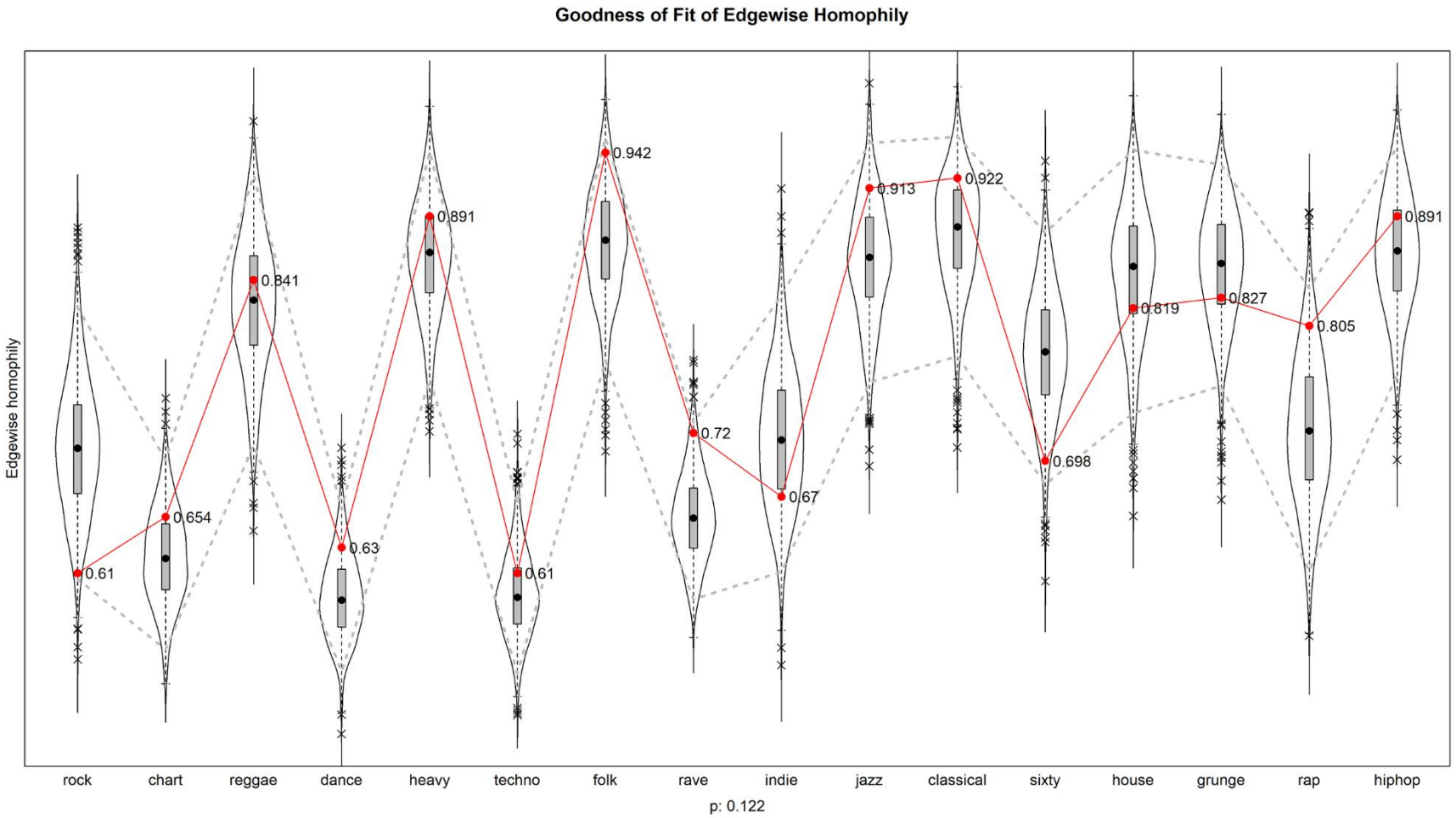


**Figure 10.** GOF testing for edgewise homophily in music style preference.

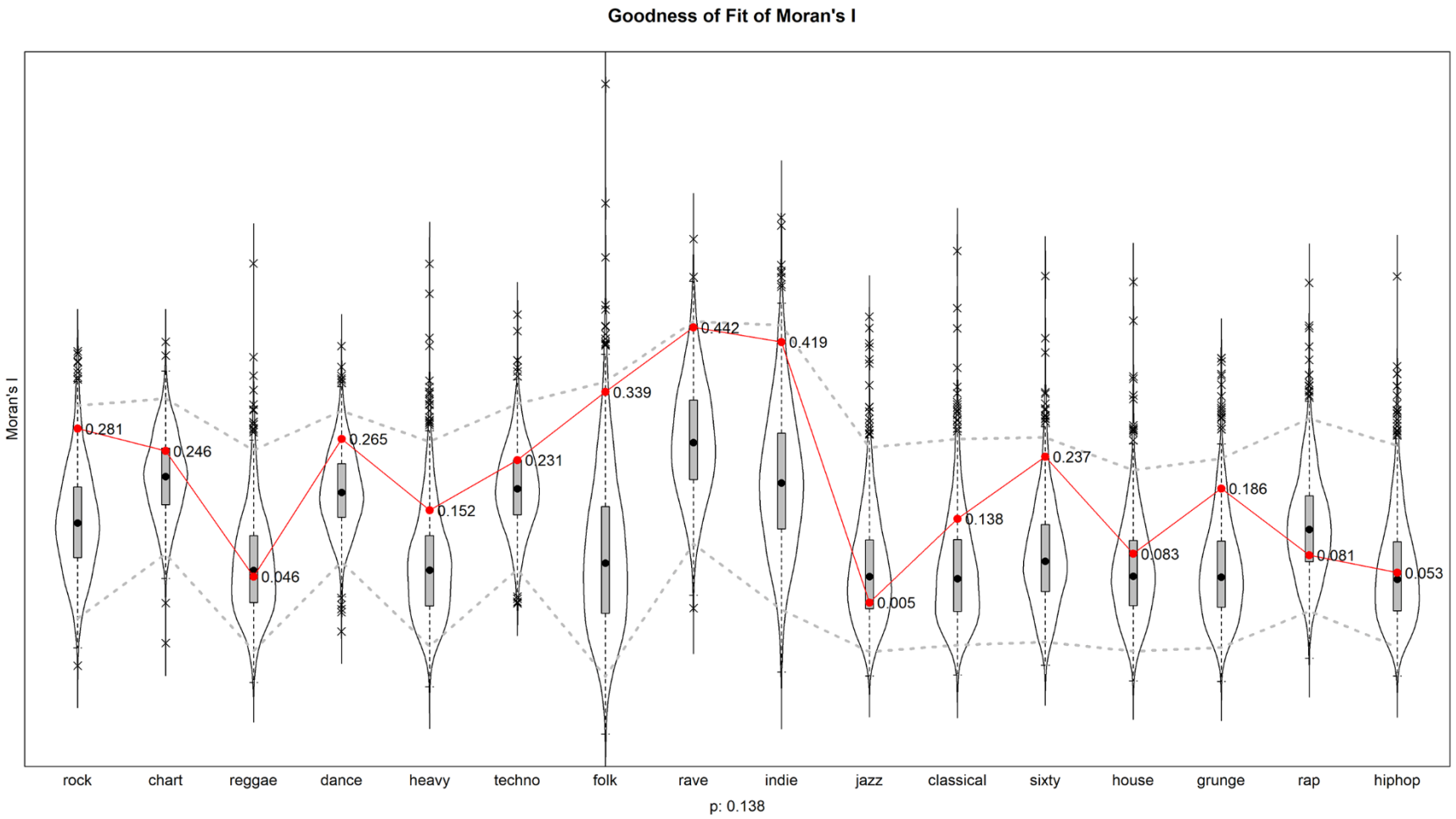


**Figure 11.** GOF testing for Moran's I in music style preference.

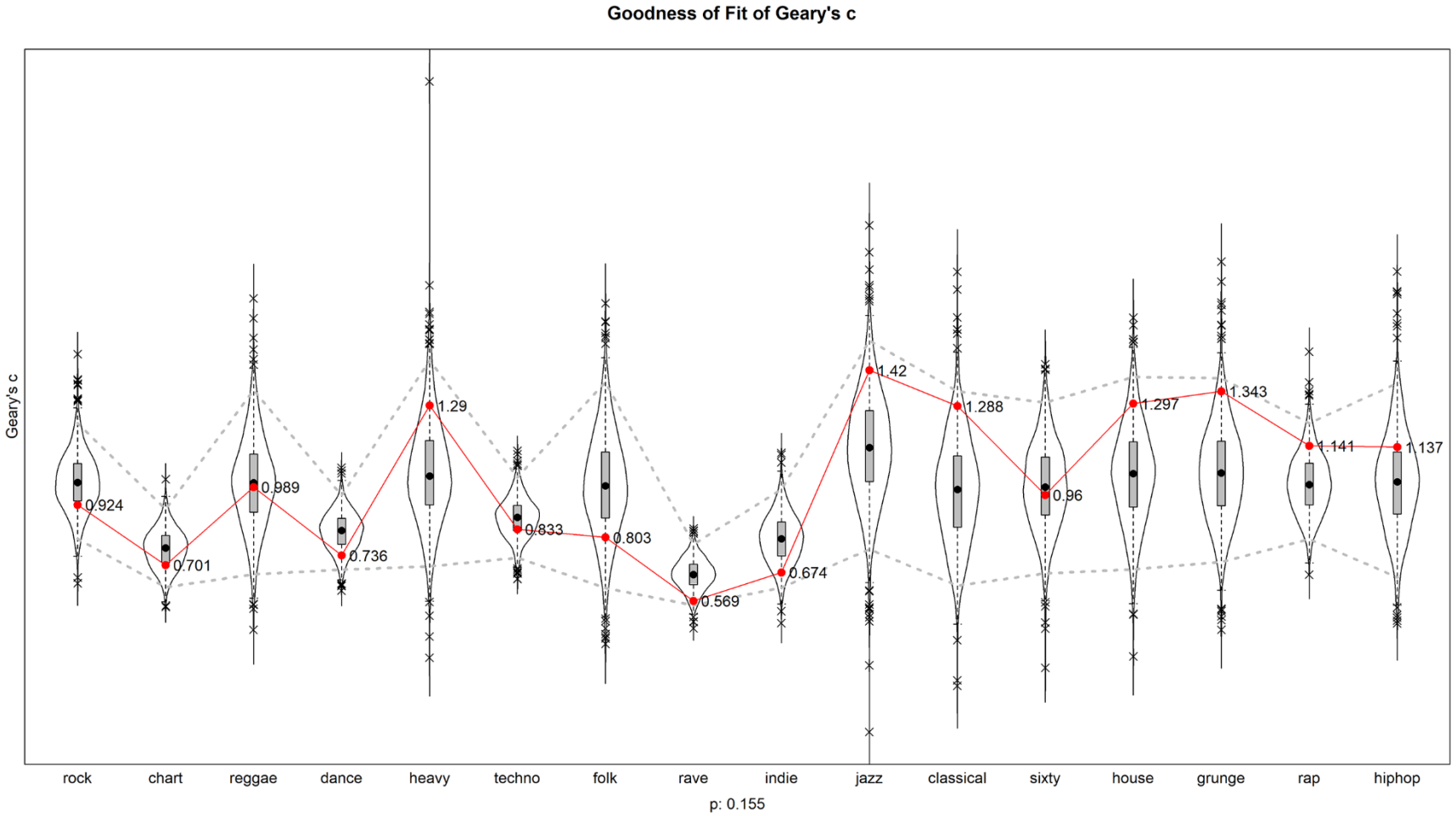


**Figure 12.** GOF testing for Geary's c in music style preference.